

Impact of Automation Support on the Conflict Resolution Task in a Human-in-the-Loop Air Traffic Control Simulation

Joey Mercer*, Ashley Gomez**, Cynthia Gabets**, Nancy Bienert**, Tamsyn Edwards**, Lynne Martin**, Vimmy Gujral**, and Jeffrey Homola*

*NASA Ames Research Center, Moffett Field, CA, USA, 94035

(e-mail: [joey.mercer, jeffrey.r.homola]@nasa.gov)

**San Jose State University / NASA Ames Research Center, Moffett Field, CA, USA, 94035

(e-mail: [ashley.n.gomez, cynthia.gabets, nancy.bienert, tamsyn.e.edwards, lynne.h.martin, vimmy.gujral]@nasa.gov)

Abstract: To determine the capabilities and limitations of human operators and automation in separation assurance roles, the second of three Human-in-the-Loop (HITL) part-task studies investigated air traffic controllers' ability to detect and resolve conflicts under varying task sets, traffic densities, and run lengths. Operations remained within a single sector, staffed by a single controller, and explored, among other things, the controller's responsibility for conflict resolution with or without their involvement in the conflict detection task. Furthermore, these conditions were examined across two different traffic densities; 1x (current-day traffic) and a 20% increase above current-day traffic levels (1.2x).

Analyses herein offer an examination of the conflict resolution strategies employed by controllers. In particular, data in the form of elapsed time between conflict detection and conflict resolution are used to assess if, and how, the controllers' involvement in the conflict detection task affected the way in which they resolved traffic conflicts.

Keywords: Human factors, air traffic control, human-in-the-loop simulation, function allocation, human-automation interaction.

1. INTRODUCTION

The transition to NextGen will likely include increasing levels of automation to help controllers perform their duties. A progression towards higher levels of automation could enable the controllers' working environment to move from tactical separation management to strategic decision-making. Such automation is envisioned to expand performance beyond today's limits by off-loading workload from controllers onto automated functions for the majority of routine operations (JPDO, 2010). However, the nature of this human-automation team is not well understood. It is still unknown exactly which tasks are best allocated to the human operator as opposed to the automation, and vice-versa. In considering this system as a whole, careful and thorough investigation is needed to better understand, not only how each team member performs in such environments, but also any associated human-automation cooperation issues.

1.1 Motivation

The motivation behind these investigations is to address a well-known problem: current-day air traffic control techniques are very labor intensive, and are limited to the amount of information controllers can process and keep in their working memory (Ericsson & Kintsch, 1995). Function allocation is but one approach to this problem, wherein

automation can take responsibility for some tasks, theoretically easing the controller's workload.

The current series of studies fall under NASA's revised function-allocation research plan, which calls for advancing our understanding of the related air-ground and human-automation issues. In particular, the Airspace Operations Laboratory (AOL) focused on the following question: "Which separation assurance functions can air traffic controllers effectively perform in future air traffic management systems?" Understanding the strengths and weaknesses of individual team members is an important aspect in determining how to distribute tasks between team members. As a first step towards gaining such insights into human-automation teaming, our approach has been to conduct part-task HITL simulations that identify the capabilities and limitations of the controller in key separation assurance tasks.

1.2 Function Allocation Research

In May of 2015, the AOL at NASA's Ames Research Center (see Prevôt, 2014) conducted the second in a series of studies that explored the capabilities and limitations of human operators with regard to the separation assurance element of air traffic control. Specifically, the research sought to better understand how best to allocate functions between controllers and automation, using the conflict-related tasks as its main focus. The general approach sought to tease apart a primary

task from related secondary tasks. While looking across varying levels of automation, the studies measured the overall impact on the performance of the primary task. Of particular interest to the second study was discovering whether removing controllers' involvement in the detection task would impact their ability to resolve conflicts.

The first study, referred to as the Human-Automation Conflict Detection study (or HACD), and the second study, referred to as the Human-Automation Conflict Resolution study (or HACR), are reported by Edwards (2016), Homola (2016), Mercer (2016a), and Mercer (2016b). However, this paper also includes, in the following section, a brief description of the HACR simulation environment, establishing the appropriate context for the later discussions.

2. METHOD

HACR examined controller performance on the conflict resolution task under different run lengths, traffic density levels, and task sets, where the group of tasks under the controller's responsibility (versus those under the automation's responsibility) defined a given task set. Although the full study featured a 5x2x2 within-subject repeated-measures design, the scope of this paper and its analyses are limited to the following two of the study's five task sets: *Conflict Resolution* and *Conflict Detection & Resolution*. This paper also examines the traffic density variable.

2.1 Conflict Resolution Condition

The *Conflict Resolution* condition's aim was to fully isolate the conflict resolution task, and in doing so, removed the controller from the conflict detection task. The study accomplished such isolation by developing a display capability that suppressed all air traffic from the radar display unless the automation (i.e., a trajectory-aided conflict probe) detected a potential conflict. Once the automation detected a conflict, the system would turn off the 'blackout' mode, and display all traffic as it normally would, albeit with the aircraft in conflict highlighted (see Figure 1). At this point, the automation's task of detecting the conflict was complete, and it was then the controller's responsibility to issue whatever control instructions they deemed appropriate. When the automation no longer detected any conflicts, the blackout mode resumed, and remained in effect until the next conflict presentation.

2.2 Conflict Detection & Resolution Condition

The Conflict Detection & Resolution condition operated much like current-day air traffic control. In addition to resolving conflicts, the controller was responsible for all conflict detection efforts, necessarily keeping constant watch over their sector's radar display, observing the progress of air traffic in and around their sector, and issuing control instructions they deemed necessary.

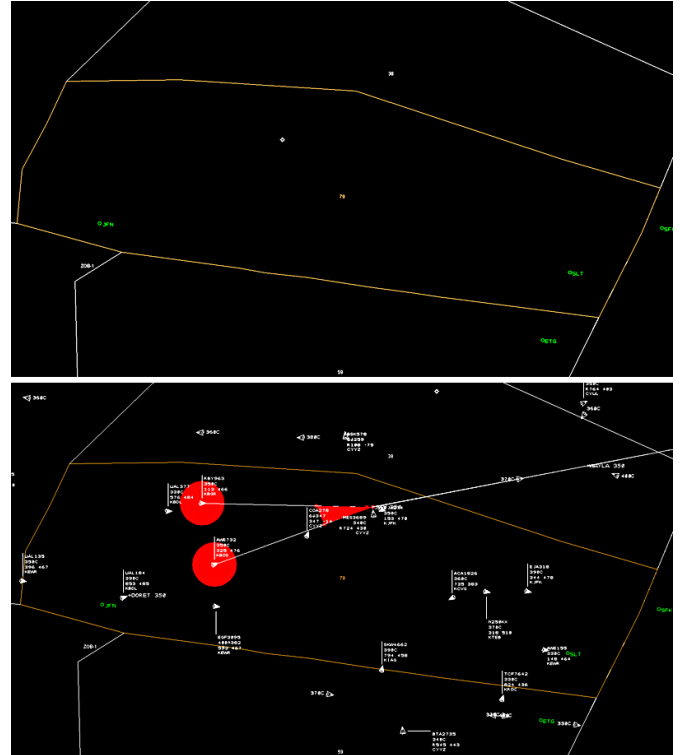


Fig. 1. Screen capture of the controller's radar display in the Conflict Resolution condition before the automation detects a conflict (top), and after the automation detects a conflict (bottom).

In order to get a clear measurement of when controllers detected a conflict, throughout the study they made keyboard entries to signal when they believed an aircraft pair to be in conflict. Without this procedure, characterizing (i.e., quantifying) the conflict resolution process across the two conditions would have been difficult. In the *Conflict Resolution* condition, measurements between an encounter's 'start' time (i.e., screen 'on' time) and the resolution time were clear. A comparable measurement from the *Conflict Detection & Resolution* condition therefore, needed a similar encounter start time, ultimately satisfied by using the time of controller's keyboard entry.

2.3 Airspace and Traffic

The airspace used during the simulation consisted of a single high-altitude sector, with a mix of overflights passing through at level altitudes, and transitioning aircraft descending to or climbing out from area airports. The scenarios progressed through a ramp-up, peak, and ramp-down phase, with each phase lasting approximately 20 minutes. Traffic levels reached 18 aircraft in the sector in the 1x traffic density, and 22 aircraft in the 1.2x density. The simulation environment also included winds for the area, which were constant-at-altitude with a nominal forecast error.

2.4 Participants

Eight retired FAA en route controllers (with an average of 24.9 years of experience among them) participated in the study, all of whom worked the same conditions. Four additional retired controllers staffing the airspace surrounding the test sector, as well as 12 pseudo pilots, worked as confederates in simulation.

Each of the eight controller participants were assigned to a specific ‘world’ that was independent of the other ‘worlds,’ but run in parallel for data-collection efficiency. To accommodate the eight parallel worlds, four physically separate rooms each housed two test sectors. To alleviate the chance of controllers being influenced by each other, the study design was such that the two controllers sharing a room never ran the same study condition at the same time. This approach helped to limit the introduction of external variables and maintained potential cross-study comparisons between HACD and HACR, since HACD also used the same parallel-worlds methodology.

2.5 Equipment

The primary simulation platform used for the study was the Multi Aircraft Control System (MACS), which, for each controller workstation, hosted an En Route Automation Modernization (ERAM) emulation on a large-format monitor (Prevôt, 2014). The controller workstation also included a specialized keyboard and trackball, similar to those used in current air traffic control facilities, as well as a custom, stand-alone voice application emulating the fielded communication system. Data recorded and collected at each workstation included aircraft flight states, operator task data and workload, automation states, and voice communications.

2.6 Training and Data Collection Schedule

The study took place over four continuous days within the same week. After an initial briefing, the remainder of the first day served to train the controllers on the study environment and procedures. The other days were devoted to the data collection effort, which produced a total of 20 runs to encompass the study’s design. The controllers completed questionnaires at the end of each run, as well as a post-simulation questionnaire. Debrief discussions provided an additional opportunity for controllers to offer feedback.

3. RESULTS

The current analyses examine the impact of the conflict detection task on the manner in which controllers resolved traffic conflicts. We theorize that it takes controllers more time to identify a resolution for a conflict when removed from the detection task, as opposed to when they are engaged in the detection task.

3.1 Resolution Response Time

The difference between the time at which the controllers issued a clearance to resolve a conflict, with the time of that conflict’s detection, yields the Resolution Response Time

measurement. In the *Conflict Detection & Resolution* (CD&R) condition, the detection time was marked when the controller made a keyboard entry to signal they believed an aircraft pair to be in conflict. In the *Conflict Resolution* (CR) condition, the detection time was marked when the automation identified an aircraft pair to be in conflict (i.e., typically when the ‘blackout’ mode turned off).

Histogram analyses arranging the resolution response times into 15-second bins revealed that, in general, the controllers were able to issue resolution maneuvers within 30 seconds of a conflict’s detection time for 49% of cases in the CR condition, but did so for 59% of cases in the CD&R condition (see Figure 2).

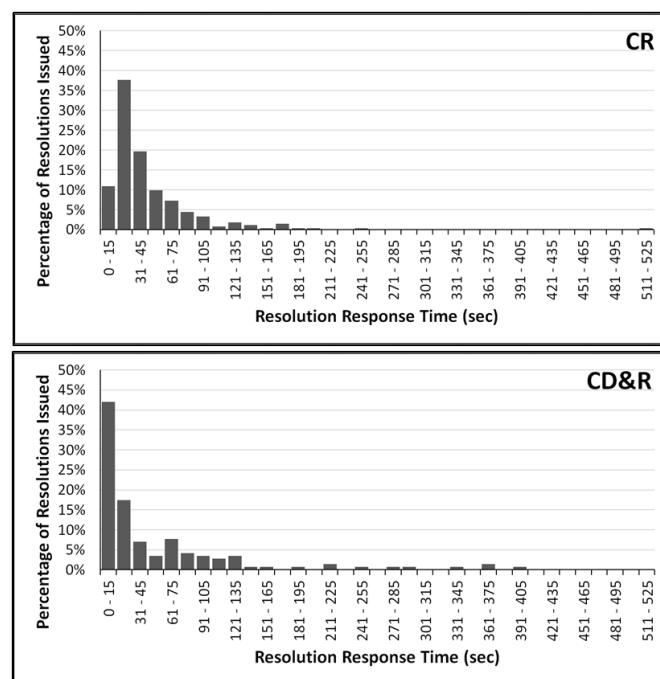


Fig. 2. Histograms showing the distribution of resolution response times for the two task sets.

After accounting for the traffic density variable, the same trend held true: the proportion of resolution maneuvers issued within 30 seconds of conflict detection were 46% and 56% for the same conditions (respectively) at the 1x traffic density, and 51% and 64% at the 1.2x density.

A clear contributor to resolution response time is the context surrounding the conflict. After identifying aircraft vertical state as a key element of a given conflict’s ‘nature’, a separate analysis categorized the resolution response time data according to the predicted vertical states of the aircraft in conflict. Two categories emerged from this analysis: conflicts where the predicted loss of separation would involve both aircraft at a constant, level altitude, and conflicts where the predicted loss of separation would involve one or both aircraft in a transitioning altitude (i.e., climbing or descending). Comparisons between the two conflict categories across each task set revealed maneuver proportions of 52% and 61% for the CR and CD&R conditions for level-

level conflicts, and proportions of 36% and 43% for the CR and CD&R conditions for non-level-level conflicts.

4. DISCUSSION

The resolution response time data provides a direct measure of how long it took the controller, from the time of a conflict's detection, to issue a resolution. In effect, this metric tells us the amount of time needed by the controller to figure out what to do, and when compared between the CR and CD&R task sets, helps to quantify the relationship between the conflict detection and conflict resolution tasks.

Not only at the broadest level, but also for comparisons that separated the effects of traffic density or conflict category, controllers were more often able to 'quickly' determine and issue a resolution maneuver in the CD&R condition than in the CR condition. These results suggest that involving the controller in the detection task helped them to instruct a resolution maneuver in less time. The data most in support of this trend came from the task set comparison at the denser traffic level. In fact, during the CD&R-1.2x condition, controllers had the highest concentration of resolution maneuvers issued within 30 seconds of conflict detection. Such evidence of a possibly clearer benefit in more complex situations suggests another interpretation: that involving the controller in the detection task helped them to instruct a resolution maneuver with perhaps less effort. This may be an indication of better situation awareness, since in the CR condition the controllers were primarily responding to highlighted conflicts, rather than actively scanning for them, as they were in the CD&R condition. The importance of situation awareness developed via active engagement in the task is well established: "Situation awareness is essential for controlling; controllers must develop and maintain an accurate mental model of the dynamic traffic situation in order to plan and respond appropriately" (Endsley & Rodgers, 1994).

In addition to comparisons between task sets, the findings of the conflict category analysis revealed a notable impact *within* each of the two task sets. For conflicts consisting of one or more transitioning aircraft, a smaller proportion of the controllers' clearances occurred within 30 seconds after the conflict's detection, as compared to the same task set's level-level conflicts. However, it is important to note that a majority of the traffic scenarios' conflicts were in the level-level category, and as such, one should use caution when interpreting these results, since this analysis examined sample sets of very different sizes.

The resolution response time metric, as collected in HACR, does have its flaws. The controllers' responses to the questionnaires uncovered the first such consideration. The controllers noted that the CR condition had more screen clutter than the CD&R condition. In typical operations, controllers constantly adjust data block positions for legibility and organization. In the CR condition's blackout mode however, that data block management task cannot occur. When the automation detected a conflict and disengaged the blackout mode, the controller may have seen a screen with overlapping and unorganized data blocks. If

such clutter disrupted the controller's efforts to address a conflict, they may have needed to adjust some of the data block positions first, potentially adding to their resolution response time. Another aspect relates to the initial conflict detection time used as the reference point for the calculation of the resolution response time data. In the CR condition, automated logging made it clear when the blackout mode was disengaged as a result of a detected conflict. By comparison, in the CD&R condition, controllers needed to make keyboard entries to signal they had detected a conflict. Such entries are naturally subject to human error, and may include artifacts of the controller forgetting to make the entry immediately at the moment they detected a conflict, potentially lowering their resolution response time. Also, the current-day coordination procedures between sectors can limit when controllers take certain actions, and were observed in HACR. More specifically, controllers are not allowed to maneuver aircraft outside of their sector without prior coordination with, and permission from, the neighboring sector. As such, the detected conflict between two aircraft on converging courses, that are both outside of the test sector at the time of detection, wasn't always resolved right away; sometimes the controllers would wait until one or both aircraft were inside their sector before issuing a maneuver.

These limitations in the data show why the study's operational environment was not intended to represent an actual concept: if such a system were to exist, it would need to address (or suffer from), each of those issues. As further evidence of the need to look at this data in isolation, rather

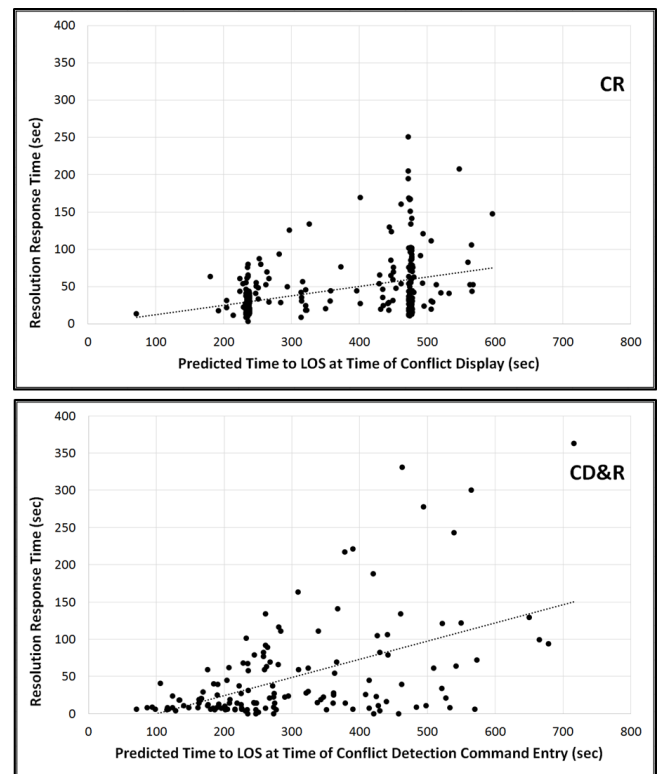


Fig. 3. Scatter plots showing the distribution of resolution response times as a function of time until predicted loss of separation (LOS), for the two task sets.

than in the context of an operational concept, consider Figure 3, which depicts the relationship between: 1) a conflict's predicted time-until-LOS at time of detection, where 'detection' was performed by the automation in the CR condition and by the controller in the CD&R condition, and 2) the resolution response time.

The scatter plots in Figure 3 show a possible shift in strategy dependent on whether controllers are, or are not, responsible for the conflict detection task. In the CR condition, there is an obvious grouping of resolutions just after the automation displayed the conflict, with very few resolutions issued with less than 200 seconds until the predicted loss of separation, whereas in the CD&R condition, there is more of a natural distribution of detection and resolution response times, with many resolutions within 200 sec of the predicted loss of separation. A review of screen recordings showed that controllers were working their resolutions into a larger plan; sometimes waiting to issue a clearance in order to facilitate a more effective series of resolutions down the road. In contrast, the automated detection of the CR condition facilitated more of a "see and fix" situation, wherein the controller was likely to be less strategic in their maneuvers, possibly gaining just enough situation awareness to resolve the conflict safely.

The apparent shift in strategy supports *the substitution myth*: a false belief that substituting the entity performing a task in order to, for example, compensate for known weaknesses, will preserve the basic system while improving its overall performance. The truth however, is that such redistributions change how each entity effectively works within the system (Dekker & Woods, 2002).

5. CONCLUSION

The results of these analyses offer evidence that, for human operators, involvement in the conflict detection task contributes to better resolution performance; when measuring performance in terms of resolution response time. The consistency of these results across all of the between-task-set comparisons adds strength to this finding. Mercer (2016b) will complement these analyses by exploring the relationship of resolution response time to the qualitative metrics of situation awareness and workload.

These results came from an environment designed to measure the operator's ability to resolve conflicts with and without a priori knowledge of the situation. This study chose not to provide the controllers with tools to help them quickly augment their understanding of the situation (e.g., an interactive trial-planner capable of displaying real-time conflict-probe feedback), in order to avoid confounding the results with specific software implementations. It is important to remember the part-task nature of this function allocation study: although the CR condition's blackout mode offered unique and interesting environments, it was never meant as an operational concept, but rather a means to better examine the relationship between detecting conflicts and resolving them.

Builders of future systems incorporating function allocation schemes - schemes in which controllers are still responsible for conflict resolution, must recognize that employing automation to 'relieve' the controller from the detection task is not without cost. These results do not claim that such allocation schemes are bad or that such costs are insurmountable: builders could point to benefits elsewhere in their system that outweigh such characteristics, and builders could provide helpful decision-support tools to the controllers. These results do show that although certain tasks can be automated, whether or not they should requires careful consideration.

The findings from this study are an initial step towards understanding the limitations of human operator performance in the air traffic control environment. Despite the shortcomings related to the primary metric, these results confirm the need for more research that can identify dependencies between other component tasks, helping to inform the proper 'placement' of automation support for effective human-automation teamwork.

REFERENCES

- Dekker, S. W. A. & Woods, D. D. (2002). MABA-MABA or Abracadabra: Progress on human automation cooperation. *Cognition, Technology and Work*, 4(4), 240-244.
- Edwards, T., Homola, J., Mercer, J., and Claudatos, L. (2016, in press). Multi-factor Interactions and the Air Traffic Controller: The Interaction of Situation Awareness and Workload in Association with Automation. *Proceedings of the 13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*.
- Endsley, M. R., & Rodgers, M. D. (1994, October). Situation awareness information requirements analysis for en route air traffic control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 38.1, 71-75.
- Ericsson, K. A., & Kintsch, W., 1995. Long term working memory. *Psychological Review*, 102, 211-245.
- Homola, J. (2016, in press). A Comparison of Conflict Detection Performance and Characteristics Between Air Traffic Controllers and a Conflict Probe in a Human-in-the-Loop Simulation. *Proceedings of the 13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*.
- Joint Planning and Development Office (2010). Concept of Operations for the Next Generation Air Transportation System. Version 3.2, JPDO, Washington, D.C.
- Mercer, J., Homola, J., Edwards, T., Gomez, A., Bienert, N., Martin, L., et. al. (2016a, in press). Assessing Operator Capabilities in a Human-Automation Team Environment during Simulated Air Traffic Control Work. *Human Factors*.
- Mercer, J., Gabets, C. Gomez, A., Edwards, T., Bienert, N., Claudatos, L., et. al. (2016b, in press). How Important is Conflict Detection to the Conflict Resolution Task? *Proceedings of the 7th International Conference on Applied Human Factors and Ergonomics, Orlando, FL, USA*.

Prevôt, T., Smith, N., Palmer, E., Callantine, T., Lee, P., Mercer, J., et al. (2014). *An Overview of Current Capabilities and Research Activities in the Airspace Operations Laboratory at NASA Ames Research Center*. AIAA-2014-2860, American Institute of Aeronautics and Astronautics, Reston, VA.